

A. Scientific and Technical Objectives

The long-term objective of our proposed studies is the development of a broad-spectrum platform for engineering complexes of metabolic enzymes in bacteria that is expected to dramatically improve the yield of virtually any biosynthetic target produced in microorganisms by accounting for enzyme organization and minimizing metabolic cross-talk. Related studies are currently funded under ONR YIP grant # N000140610565 and focus on the production of 1,2-propylene glycol (1,2-PG). The objective of this particular application is to improve microbial synthesis of D-1,2,4-butanetriol (D-BT), a precursor to the energetic material D-butanetriol trinitrate (D-BTTN), using our technology for engineering metabolic complexes in living cells. We have proposed the following specific aims:

Specific Aim 1: Co-localization of D-BT enzymes into functional metabolic complexes. Under this aim, we will assemble the D-BT pathway enzymes into functional complexes using covalent and non-covalent methods of assembly/crosslinking developed here and in conjunction with studies under N000140610565. We will also explore the use of computational tools as a means to design protein-protein interactions *de novo*.

Specific Aim 2: Enable combinatorial engineering of metabolic complexes via metabolite sensors. We will engineer a collection of protein-based switches that are capable of dynamically responding to our desired end-product D-BT over a broad concentration range. Such sensors will enable fine-tuning (e.g., laboratory evolution) of our engineered D-BT channels.

B. Approach

Our approach to improving D-BT biosynthesis is to sequentially tether enzymes together to form synthetic metabolic complexes inside living bacterial cells. The rationale for building synthetic complexes is that D-BT titers will be increased by accounting for pathway enzyme organization in a manner that (1) eliminates intracellular enzyme/substrate diffusion barriers and thus increases catalytic efficiency by co-localizing enzymes to discrete locations in the cells; (2) enables control over the molecular ratios of pathway enzymes; and (3) minimizes metabolic cross-talk and side reactions by focusing metabolic flux towards the production of D-BT and away from known side products. To accomplish our objective, sequential tethering of D-BT pathway enzymes into either static or dynamic channels will be performed by: (1) construction of fusion proteins (static); (2) post-translational protein assembly using covalent (static) or non-covalent bonds (dynamic); and (3) enzyme co-immobilization to an intracellular scaffold (dynamic). Once we have assembled the D-BT pathway in a complex inside living cells, we will quantify the formation of D-BT as well as all unwanted side products to determine the net effect of different enzyme assembly strategies. While our approach is largely experimental in nature, we will also explore the use of computational studies to design alternative assembly strategies whereby interactions between metabolic enzymes is designed *de novo*. Finally, in parallel to the assembly of the D-BT metabolic enzymes, we will use develop a set of D-BT sensors based on a chemical genetic reporter of protein stability. These reporters will be useful for direct monitoring of intracellular D-BT titers in living cells and are expected to open the door to laboratory evolution of our engineered metabolic complexes.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2008		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE 2008 Annual Report				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Cornell University				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 10	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

C. Concise Accomplishments

The major accomplishments to date include:

(1) We applied our previously developed spatial stochastic model (Conrado *et al.*, 2007) to D-BT biosynthesis and demonstrated that compartmentalization of the four pathway enzymes comprising an engineered pathway for D-BT biosynthesis leads to greatly improved kinetic properties for the pathway enzymes. A manuscript detailing these findings is in preparation.

(2) We have developed a chemical genetic reporter of protein stability that enables intracellular sensing of small compounds (Haitjema *et al.*, 2008). The significance of this tool is that it provides a fluorescence-based reporter of metabolite levels thereby opening the door to laboratory evolution of our metabolic D-BT assemblies.

(3) We have engineered two novel assays for studying and engineering intracellular protein-protein interactions (Contreras-Martinez & DeLisa, 2007, Waraho & DeLisa, 2008). These assays provide a powerful toolkit for fine-tuning the affinity between the D-BT metabolic enzymes.

(4) We have developed computational tools for the study and design of protein-protein interactions (Contreras Martinez *et al.*, 2008). These tools should be useful in the de novo design of protein-protein interfaces that promote assembly of D-BT metabolic enzymes.

D. Expanded Accomplishments

D1. Simulation of D-BT enzyme co-localization. Our first major accomplishment has been the demonstration that enzyme co-localization has a measurable effect on D-BT titers in *Escherichia coli*. This was demonstrated via computer simulation in collaboration with Dr. Jeffrey Varner (Cornell University). Progress was made on these studies in year 1 of this grant; during year 2 we have completed these studies and a manuscript is in preparation. Specifically, we developed a spatial stochastic model of *E. coli* central carbon metabolism using the Next Subvolume Method, an efficient implementation of the Gillespie direct method of stochastic simulation. Using this model, we previously demonstrated that synthetic metabolic complexes locally improve the catalytic efficiency of an enzyme assembly for 1,2-propanediol biosynthesis compared to the unassembled case (Conrado *et al.*, 2007). Under this award, we have applied our spatial stochastic model to D-BT biosynthesis and demonstrated that discrete co-localization of the four pathway enzymes comprising an engineered pathway for D-BT biosynthesis leads to improved kinetic properties for the pathway enzymes, especially when substrate diffusivities are low. Our results suggest that enzyme co-localization is a powerful approach for improving the catalytic turnover of a channeled carbon substrate and should be particularly useful when applied to synthetic metabolic pathways that suffer from poor translation efficiency, are present in highly variable copy numbers, and have low turnover for new substrates. Furthermore, this approach represents a generic modeling framework for simultaneously analyzing spatial and stochastic events in cellular metabolism and should enable quantitative evaluation of the effect of enzyme compartmentalization on virtually any recombinant pathway.

D2. Chemical genetic control of protein stability results in an intracellular sensor of small molecules.

Since there are currently no generic reporters for intracellular metabolites, we sought to develop a tool for sensing such compounds. Our approach was to develop a protein conformational switch comprised of an unstable domain and a reporter protein (Fig. 1). The unstable domain was selected such that introduction of a small molecule ligand that stabilizes the domain would restore stability of the entire fusion and thus lead to measurable activity of the reporter protein. The reporter for our switch was chosen as the green fluorescent protein (GFP) so that upon introduction of small molecules that stabilize the unstable domain, a large increase in cell fluorescence would result (Fig. 2a). For the unstable domain, we chose the TraR transcriptional activator from *Agrobacterium tumefaciens* (Fig. 2a). In the absence of its natural ligand, the freely diffusible quorum signaling molecule 3-oxooctanoyl-L-homoserine lactone (OHHL), the TraR protein is a monomer that is highly unstable in the cytoplasm of *Agrobacterium tumefaciens* and *E. coli* (Zhu & Winans, 2001). However, upon binding of OHHL, TraR forms an extremely stable dimer (Vannini *et al.*, 2002). We have observed this same OHHL-dependent stability with an engineered TraR-GFP fusion protein. That is, in the absence of OHHL, TraR-GFP is highly unstable and cells expressing the fusion are relatively non-fluorescent (Fig. 2b). However, upon addition of OHHL, the TraR-GFP protein is stabilized (presumably in a dimeric conformation) and the cells become highly fluorescent in an OHHL dose-dependent fashion (Fig. 2b). This work was recently submitted for publication (Haitjema *et al.*, 2008). We are now exploring the further engineering of this GFP-TraR conformational switch for sensing molecules other than OHHL. We expect that a collection of small molecule switches can be created using the GFP-TraR backbone, simply by the application of protein design and/or laboratory evolution to change the substrate specificity of TraR from OHHL to other compounds of interest such as D-BT.

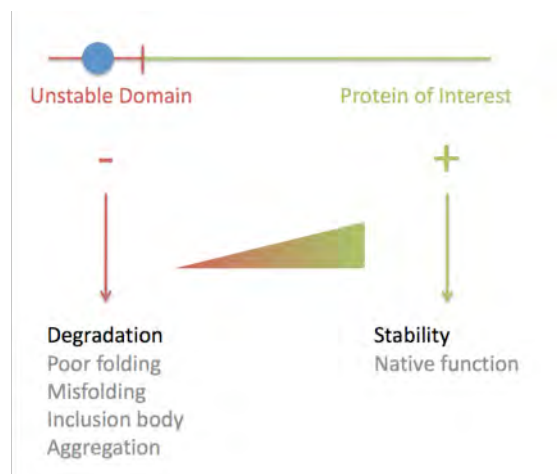


Figure 1. Chemical genetic control of protein stability.

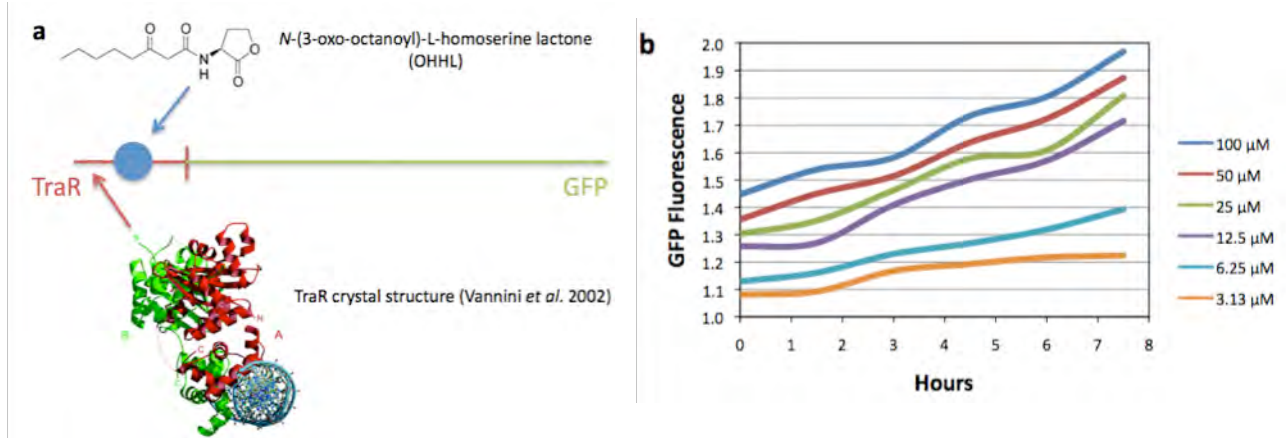


Figure 2. (a) TraR-GFP fusion as a reporter of small molecules in the cytoplasm of *E. coli* cells. (b) Dose-dependent response of the TraR-GFP-expressing *E. coli* to various concentrations of OHHL added exogenously to the growth medium. Cells were grown in 96-well plates and assayed using a fluorescent plate reader.

D3. Genetic assays for engineering protein-protein interactions. To characterize and engineer the affinity between metabolic pathway enzymes, we have developed two high-throughput assays for monitoring protein-protein interactions in the cytoplasm of *E. coli* (Contreras-Martinez & DeLisa, 2007, Waraho & DeLisa, 2008). These are described below.

D3.1 A versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism (Waraho & DeLisa, 2008). We have developed a reliable genetic selection strategy for isolating interacting proteins based on the 'hitchhiker' mechanism of the *E. coli* twin-arginine translocation (Tat) pathway. This method, designated FLI-TRAP (Functional Ligand-binding Identification by Tat-based Recognition of Associating Proteins), is based upon the Tat system's unique ability to efficiently co-translocate non-covalent complexes of two folded polypeptides (see Fig. 3). In the FLI-TRAP assay, one polypeptide is a fusion between an N-terminal Tat signal peptide (ssTorA) and the protein to be screened for interactions and the second is a fusion between a known or putative partner protein and mature TEM-1 β -lactamase (Bla). Using a series of c-Jun and c-Fos leucine zipper (JunLZ and FosLZ) variants of known affinities, we observed that only those chimeras that expressed well and interacted strongly in the cytoplasm were able to co-translocate Bla into the periplasm and confer β -lactam antibiotic resistance to cells. The utility of this assay was then demonstrated through: (1) random library selection of amino acid substitutions that restored heterodimerization to a non-interacting FosLZ variant; and (2) creation of artificial interaction interfaces between a Tat-targeted soluble scaffold (e.g., green fluorescent protein) and a cognate binding target fused to Bla. Since Tat substrates must be correctly folded prior to transport, FLI-TRAP favors the identification of soluble, non-aggregating, protease-resistant protein pairs and thus provides a powerful tool for routine selection of interacting partners (e.g., antibody-antigen) without the need for purification or immobilization of the binding target. We are now using this assay to engineer interactions between metabolic enzymes, where 2 enzymes of interest from the D-BT pathway are introduced into the X and Y positions of the reporter system. This can be used to interrogate any mode of non-covalent interaction between the enzymes of interest.

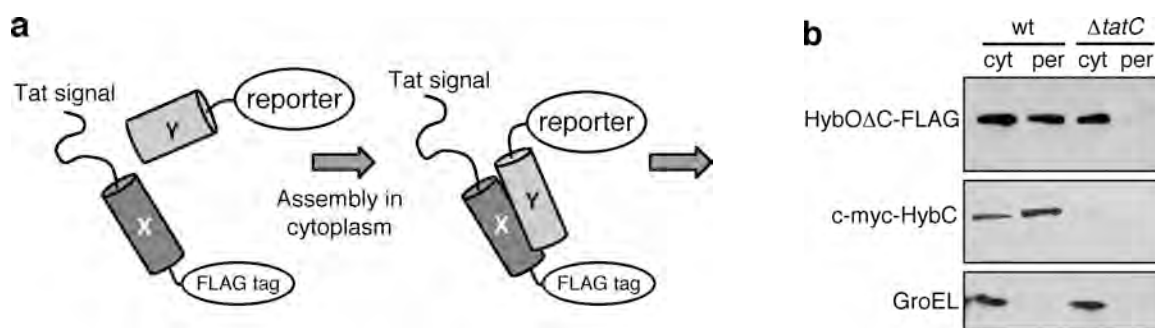


Figure 3. Tat-mediated hitchhiker export in bacteria. (a) Schematic of engineered assay for co-translocation of interacting pairs via the Tat pathway. In this study, the Tat signal peptide chosen was ssTorA, the reporter was either the c-myc epitope tag or β -lactamase and X and Y were interacting domains or entire protein pairs. (b) Western blot analysis of cytoplasmic (cyt) and periplasmic (per) fractions from cells co-expressing HybOΔC and c-myc-HybC, where each was detected using an anti-FLAG and anti-c-myc antibody, respectively. The cytoplasmic chaperone GroEL was detected using an anti-GroEL antibody and served as a fractionation marker.

D4. Computational approaches to protein-protein interactions. To complement our experimental strategies to engineering protein-protein interactions, we have also performed computational analysis to gain further insight into the design of *de novo* interfaces between two proteins of interest. These efforts are summarized below.

D4.1 In silico protein fragmentation reveals the importance of critical nuclei in domain reassembly (Contreras Martinez et al., 2008). Protein complementation assays (PCAs) based on split protein fragments have become powerful tools that facilitate the study and engineering of intracellular protein-protein interactions. These assays are based on the observation that a given protein can be split into two inactive fragments and these fragments can reassemble into the original properly folded and functional structure. However, one experimentally observed limitation of PCA systems is that the folding of a protein from its fragments is dramatically slower relative to that of the unsplit parent protein. This is due in part to a poor understanding of how PCA design parameters such as split site position in the primary sequence and size of the resulting fragments contribute to the efficiency of protein reassembly. We used a minimalist on-lattice model to analyze how the dynamics of the reassembly process for two model proteins was affected by the location of the split site (see Fig. 5). Our results demonstrate that the balanced distribution of the “folding nucleus”, a subset of residues that are critical to the formation of the transition state leading to productive folding, between protein fragments is key

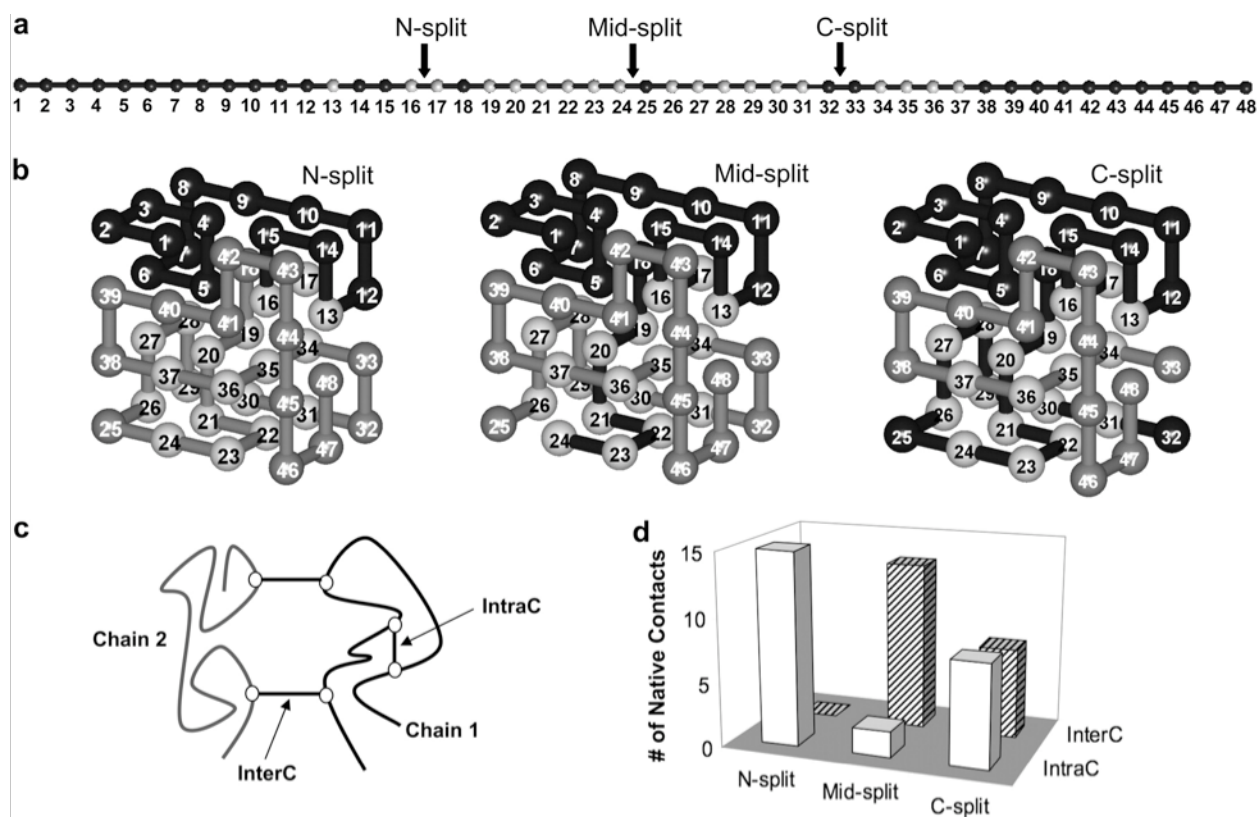


Figure 5. Simulation of protein assembly systems. (a) Linear representation of the parent 48-mer sequence. Fragmentation sites for N-, Mid- and C-split systems are indicated. Residues comprising the folding nucleus at the center of the folded structure are shown in light gray. (b) Schematic representation of folded structures for split systems. Amino acids or connecting bonds shown in black correspond to N-terminal fragment (chain 1); amino acids or connecting bonds shown in dark gray correspond to those in the C-terminal fragment (chain 2); amino acids shown in light gray correspond to the folding nucleus. (c) Schematic representation of *inter*-chain (InterC) contacts that involve interacting residues from both chains, and *intra*-chain (IntraC) contacts that involve interacting residues within the same chain. (d) InterC and IntraC contacts formed by critical core residues upon protein fragmentation. Most probable native contacts found in the transition state (TS) ensemble (i.e., the folding core/nuclei) for folding of the 48-mer sequence at $T_f = 0.27$.

to their reassembly. In addition, this worked revealed that the association and formation of a large interface between two monomers (that remain largely unfolded by themselves) are key prerequisites for the concurrent folding of the two proteins as one stable complex. Collectively, our studies in this area are providing new fundamental insights into the structural organization of metabolic machinery within living cells and should be useful in the future design of intracellular protein complexes for use in varied applications.

D4.2 Computational design of protein-protein interfaces using RosettaDock. We have begun a collaboration with Dr. David Baker (University of Washington) to explore computational design of protein-protein interactions *de novo*. To accomplish this, we are using RosettaDock, a command line application that is used to determine the conformation of interacting proteins. The two main modes of this application are perturbation mode and *ab initio*. *Ab initio* mode will fully randomize the position of both partners and attempt to locate a binding interface, while perturbation mode makes small adjustments to the starting orientation of the partners. Every time the partners are moved, the free energy is calculated in addition to root mean squared distance relative to the starting structure. This information is used to generate a plot that ultimately shows if the orientation of the partners is correct. If the supplied file has the partners aligned properly, there will be a sharp decrease in energy as the rmsd approaches zero. This is known as an energy funnel. Because of the energy scoring algorithm that RosettaDock uses, perturbation mode is much more reliable than *ab initio*. This means that the orientation of the interaction must be roughly predicted by the user to get the best results.

As a model system, we have employed RosettaDock to explore the interaction between two known interacting proteins, namely the ubiquitin ligase MDM2 and one of its known interacting partners, the tumor suppressor protein p53. Based on a molecular-level understanding of the MDM2-p53 interaction, we are trying to redesign the MDM2 ubiquitin ligase for recognizing a non-native target such as the CDK2 protein. This is roughly equivalent to the problem we are addressing with the D-BT metabolic enzymes – how do we create a binding interface between

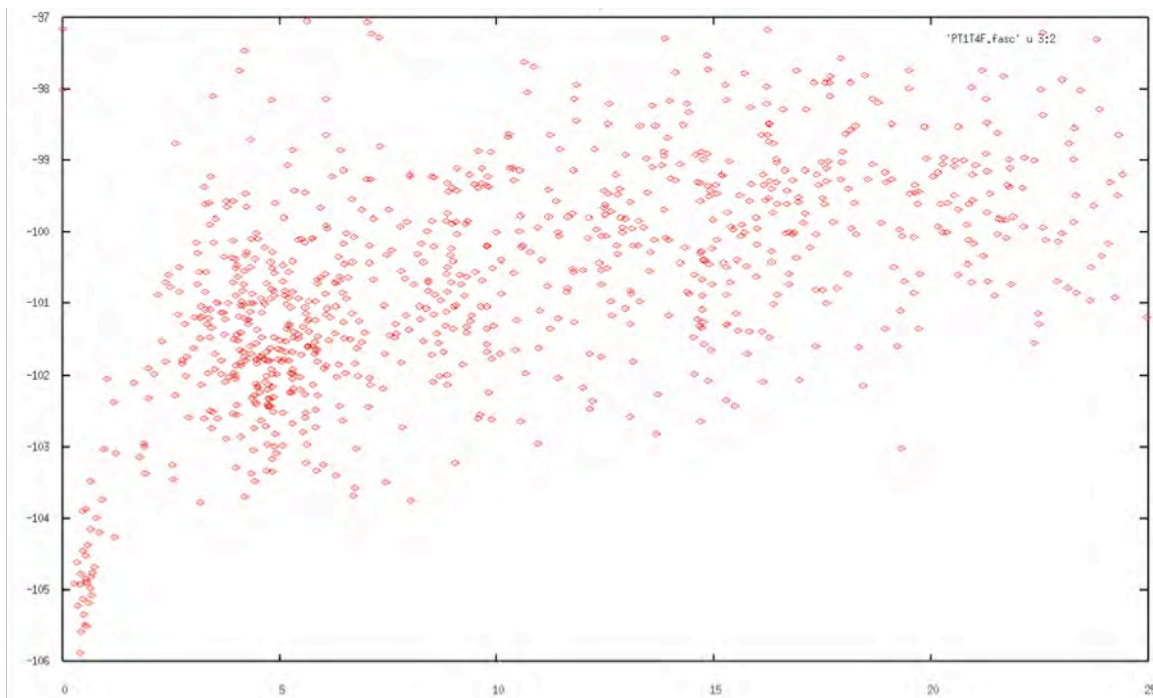


Figure 6. The above figure shows ideal output from a docking calculation. The x-axis is root mean squared distance (rmsd) from the starting structure, and the y-axis is energy. Lower energy indicates a more favorable interaction between the docking partners. The figure shows a steep drop in energy as rmsd approaches zero indicating that the initial input is the most energetically favorable orientation of the partners.

two proteins that do not natively interact? To accomplish this, we have employed several interrelated computational approaches: (1) protein visualization (MacPyMOL) to view the natural MDM2-p53 complex and identify the binding interface; (2) protein docking (RosettaDock) (see Fig. 6) and computational mutagenesis techniques (RosettaDesign) to determine the mutations necessary to successfully redesign the MDM2 binding interface for a nonnative target such as CDK2. Once successful *in silico* retargeting is completed, we will experimentally validate whether our engineered MDM2 proteins can bind CDK2. If successful, this same technique could be utilized to design putative interfaces between 2 metabolic enzymes from the D-BT pathway. This work was recently presented at the AIChE Northeast Regional Conference held at MIT in April 2008 (Weitzner *et al.*, 2008).

E. Work Plan

Frost and coworkers recently demonstrated a direct fermentation route to D-BT in engineered *E. coli*, however, titers are limited by: (1) the existence of three competing pathways that yield unwanted side-products and divert carbon flux from D-BT; and (2) the use of three heterologous enzymes that do not normally co-exist or interact in nature. Thus, the workplan moving forward will focus on the following key issues:

(1) Now that we have completed an MTA with Draths Corporation and are in possession of the D-BT pathway enzymes (see section F below), we will apply several of our experimental strategies for enzyme assembly to the D-BT enzymes in a manner identical to our metabolic engineering work on 1,2-propanediol (see ONR YIP grant # N000140610565). The enzymes for D-BT biosynthesis are: *Caulobacter crescentus* D-xylose dehydrogenase (Cc-Xdh), *E. coli* D-xylonate dehydratase (Ec-YjhG/YagF), *Pseudomonas putida* benzoylformate decarboxylase (Pp-MdIC) and *E. coli* alcohol dehydrogenase (Ec-AdhP). We will explore several different methods for the assembly of these enzymes. One approach will be covalent tethering of pathway enzymes via (i) translational fusion; or (ii) post-translation cross-linking by transglutaminase. An alternative approach will be the use of “protein interacting domains” whereby dimerization domains from eukaryotic proteins (e.g., jun/fos) will be fused onto each pathway enzyme, thereby creating artificial interaction domains that will catalyze non-covalent enzyme assembly. This will allow us to explore the extent to which the nature of the assembly (i.e., covalent bonds vs. non-covalent bonds) affects the efficiency of the metabolic conversion (i.e., D-BT production). We will initially focus on the assembly of just two enzymes, namely Pp-MdIC and Ec-AdhP, because these two enzymes are at the crucial branch point where side reactions occur. If successful, we will then explore the assembly of three- and four-enzyme assemblies.

(2) Our results to date have clearly demonstrated that the TraR-GFP fusion has an exquisite ability to be stabilized by the binding of a small molecule ligand and thus “sense” the presence of extremely small compounds. Moving forward, we will focus on the development of TraR-GFP sensors that respond to an array of different compounds. This will entail the use of rational design and laboratory evolution in order to alter the substrate specificity of TraR-GFP. We will initially focus on the substrate D-BT.

(3) We will also make use of our two protein-protein interaction assays to begin engineering high-affinity partnerships between D-BT pathway enzymes. These assays will allow us to screen directly for factors/mutations that promote strong non-covalent interaction between any 2 D-BT metabolic enzymes of interest. Any useful information that emerges as a result of our computational studies (see section D4 above) can be characterized experimentally using either of these two assays.

(4) To fully realize the benefit of metabolic channeling, we must maximize the flow of D-BT precursors into the assembly, i.e., we must embed the D-BT assembly into a metabolic strain background that has been engineered to take advantage of the benefits offered by channeling. To design optimal precursor flux to the D-BT channel, we will develop new network design tools with assistance from our collaborator Dr. Jeffrey Varner (Cornell University) that can be used to computationally develop metabolic architectures that take full advantage of engineered assemblies. Our strategy will use error tolerant kinetic models, based upon the cybernetic modeling paradigm, in conjunction with a non-linear programming procedure inspired by OPTKNOCK to calculate optimal network configurations that maximize flux to D-BT assemblies while simultaneously satisfying design and process constraints.

F. Major Problems/Issues. The only significant problem has been obtaining the necessary enzyme sequences for D-BT biosynthesis in *E. coli* from Dr. John Frost. The primary reason for this has been caused by John's departure from Michigan State University to form a small start-up company. The company, Draths Corporation, was not in a position to complete MTAs and thus we were unable to obtain the necessary materials to begin assembling the D-BT enzymes. However, we recently completed an MTA arrangement with assistance from John Raley at Draths Corporation and, as of a week ago, the enzymes needed are now in our possession. Thus, we do not see any further impediments to completing the work plan as outlined above.

G. Technology Transfer. We have submitted one new patent application and converted an earlier provisional patent to a full patent. These are listed below.

1. Haitjema, C., Fisher, A.C. and DeLisa, M.P. "Compositions and methods for controlling protein stability in bacteria with small molecules" invention disclosed.
2. Contreras Martinez, L. and DeLisa, M.P. "Protein discovery using intracellular ribosome display" (Application number PCT/US2008/071747)

References Cited

- Conrado, R. J., T. J. Mansell, J. D. Varner & M. P. DeLisa, (2007) Stochastic reaction-diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. *Metab Eng* **9**: 355-363.
- Contreras Martinez, L. M., E. E. Borrero Quintana, F. A. Escobedo & M. P. DeLisa, (2008) In silico protein fragmentation reveals the importance of critical nuclei on domain reassembly. *Biophys J* **94**: 1575-1588.
- Contreras-Martinez, L. M. & M. P. DeLisa, (2007) Intracellular ribosome display via SecM translation arrest as a selection for antibodies with enhanced cytosolic stability. *J Mol Biol* **372**: 513-524.
- Haitjema, C., A. C. Fisher & M. P. DeLisa, (2008) Chemical genetic control of protein stability results in an intracellular sensor of small molecules. *J Amer Chem Soc* (submitted).
- Vannini, A., C. Volpari, C. Gargioli, E. Muraglia, R. Cortese, R. De Francesco, P. Neddermann & S. D. Marco, (2002) The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J* **21**: 4393-4401.
- Waraho, D. & M. P. DeLisa, (2008) A versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proc Natl Acad Sci U S A* (in revision).
- Weitzner, B., J. D. Varner & M. P. DeLisa, (2008) Computational retargeting of an E3 ubiquitin ligase. In: AIChE Regional Meeting. MIT, pp.
- Zhu, J. & S. C. Winans, (2001) The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. *Proc Natl Acad Sci U S A* **98**: 1507-1512.